

WHAT IS CLAIMED IS:

1 1. A method of detecting similarity between protein sequences comprising
2 comparing a first disulfide signature to a second disulfide signature, each disulfide signature
3 being characteristic of a corresponding protein sequence.

1 2. The method of claim 1, wherein each disulfide signature describes a disulfide
2 topology of the corresponding protein sequence.

1 3. The method of claim 1, wherein each disulfide signature includes the number
2 of residues between a pair of cysteines joined by a disulfide bridge, and the number of
3 residues between the first cysteine of each disulfide bridge and the first cysteine of the next
4 disulfide bridge in the corresponding protein sequence.

1 4. The method of claim 3, wherein each disulfide signature includes the number
2 of residues between each pair of cysteines joined by a disulfide bridge, and the number of
3 residues between the first cysteine of each disulfide bridge and the first cysteine of the next
4 disulfide bridge in the corresponding protein sequence, for each disulfide bridge in the
5 corresponding protein sequence.

1 5. The method of claim 1, wherein comparing includes calculating a measure of
2 similarity between the first disulfide signature and the second disulfide signature.

1 6. The method of claim 5, wherein comparing further includes calculating a
2 measure of statistical relevance for the measure of similarity between the first disulfide
3 signature and the second disulfide signature.

1 7. The method of claim 1, wherein comparing includes searching a database
2 including a plurality of disulfide signatures, each disulfide signature of the database
3 characteristic of a corresponding protein sequence.

1 8. The method of claim 7, wherein comparing includes calculating a measure of
2 similarity between the first disulfide signature and each of a plurality of disulfide signatures
3 of the database.

1 9. The method of claim 7, wherein searching the database includes searching
2 with a subpattern of the first disulfide signature.

1 10. The method of claim 9, wherein the subpattern is generated by calculating the
2 disulfide signature that results when one or more disulfide bridges is removed from the
3 protein sequence corresponding to the first disulfide signature.

1 11. The method of claim 7, wherein at least one disulfide signature in the database
2 is associated with a sequence identifier.

1 12. The method of claim 7, wherein at least one disulfide signature in the database
2 is associated with a domain identifier.

1 13. The method of claim 7, further comprising clustering disulfide signatures of
2 the database.

1 14. The method of claim 13, wherein clustering includes grouping disulfide
2 signatures by number of disulfide bridges.

1 15. The method of claim 13, wherein clustering includes grouping disulfide
2 signatures by disulfide topology.

1 16. The method of claim 13, wherein clustering includes calculating a measure of
2 similarity between disulfide signatures and grouping based on the measure of similarity.

1 17. A method of detecting similarity between protein sequences comprising:
2 generating a database including a plurality of disulfide signatures, each disulfide
3 signature being characteristic of a corresponding protein sequence; and
4 comparing a first disulfide signature corresponding to a protein sequence to at least
5 one disulfide signature of the database.

1 18. The method of claim 17, wherein each disulfide signature describes a disulfide
2 topology of the corresponding protein sequence.

1 19. The method of claim 18, wherein each disulfide signature includes the number
2 of residues between a pair of cysteines joined by a disulfide bridge, and the number of

3 residues between the first cysteine of each disulfide bridge and the first cysteine of the next
4 disulfide bridge in the corresponding protein sequence.

1 20. The method of claim 19, wherein each disulfide signature includes the number
2 of residues between each pair of cysteines joined by a disulfide bridge, and the number of
3 residues between the first cysteine of each disulfide bridge and the first cysteine of the next
4 disulfide bridge in the corresponding protein sequence, for each disulfide bridge in the
5 corresponding protein sequence.

1 21. The method of claim 17, wherein generating the database includes identifying
2 a disulfide bridge by protein sequence homology or protein structure homology.

1 22. The method of claim 17, wherein generating the database includes calculating
2 a disulfide signature for a protein sequence.

1 23. The method of claim 17, wherein comparing includes calculating a measure of
2 similarity between the first disulfide signature and the disulfide signature of the database.

1 24. The method of claim 23, wherein comparing further includes calculating a
2 measure of statistical relevance for the measure of similarity between the first disulfide
3 signature and the disulfide signature of the database.

1 25. The method of claim 17, wherein comparing includes comparing a subpattern
2 of the first disulfide signature to at least one disulfide signature of the database.

1 26. The method of claim 25, wherein the subpattern is generated by calculating
2 the disulfide signature that results when one or more disulfide bridges is removed from the
3 corresponding protein sequence.

1 27. The method of claim 17, wherein at least one disulfide signature of the
2 database is associated with a sequence identifier.

1 28. The method of claim 17, wherein at least one disulfide signature of the
2 database is associated with a domain identifier.

1 29. The method of claim 18, further comprising clustering the disulfide signatures
2 of the database.

1 30. The method of claim 29, wherein clustering includes grouping disulfide
2 signatures by number of disulfide bridges.

1 31. The method of claim 29, wherein clustering includes grouping disulfide
2 signatures by disulfide topology.

1 32. The method of claim 29, wherein clustering includes calculating a measure of
2 similarity between at least one pair of disulfide signatures and grouping based on the measure
3 of similarity.

1 33. A method of detecting similarity between protein sequences comprising
2 generating a database including a plurality of disulfide signatures, each disulfide signature
3 being characteristic of a corresponding protein sequence.

1 34. The method of claim 33, wherein each disulfide signature describes a disulfide
2 topology of the corresponding protein sequence.

1 35. The method of claim 34, wherein each disulfide signature includes the number
2 of residues between a pair of cysteines joined by a disulfide bridge, and the number of
3 residues between the first cysteine of each disulfide bridge and the first cysteine of the next
4 disulfide bridge in the corresponding protein sequence.

1 36. The method of claim 35, wherein each disulfide signature includes the number
2 of residues between each pair of cysteines joined by a disulfide bridge, and the number of
3 residues between the first cysteine of each disulfide bridge and the first cysteine of the next
4 disulfide bridge in the corresponding protein sequence, for each disulfide bridge in the
5 corresponding protein sequence.

1 37. The method of claim 33, wherein generating the database includes identifying
2 a disulfide bridge by protein sequence homology or protein structure homology.

1 38. The method of claim 33, wherein generating the database includes calculating
2 a disulfide signature for a protein sequence.

1 39. The method of claim 38, wherein calculating the disulfide signature includes
2 determining the number of residues between a pair of cysteines joined by a disulfide bridge
3 in the protein sequence.

1 40. The method of claim 38, wherein calculating the disulfide signature includes
2 determining the number of residues between the first cysteine of each disulfide bridge and
3 the first cysteine of the next disulfide bridge in the protein sequence.

1 41. A computer program for detecting similarity between protein sequences, the
2 computer program comprising instructions for causing a computer system to compare a first
3 disulfide signature to a second disulfide signature, each disulfide signature being
4 characteristic of a corresponding protein sequence.

1 42. The computer program of claim 41, wherein each disulfide signature includes
2 the number of residues between a pair of cysteines joined by a disulfide bridge, and the
3 number of residues between the first cysteine of each disulfide bridge and the first cysteine of
4 the next disulfide bridge in the corresponding protein sequence.

1 43. The computer program of claim 42, wherein each disulfide signature includes
2 the number of residues between each pair of cysteines joined by a disulfide bridge, and the
3 number of residues between the first cysteine of each disulfide bridge and the first cysteine of
4 the next disulfide bridge in the corresponding protein sequence, for each disulfide bridge in
5 the corresponding protein sequence.

1 44. The computer program of claim 41, wherein comparing includes calculating a
2 measure of similarity between the first disulfide signature and the second disulfide signature.

1 45. The computer program of claim 44, wherein comparing further includes
2 calculating a measure of statistical relevance for the measure of similarity between the first
3 disulfide signature and the second disulfide signature.

1 46. The computer program of claim 41, wherein comparing includes searching a
2 database including a plurality of disulfide signatures, each disulfide signature of the database
3 characteristic of a corresponding protein sequence.

1 47. The computer program of claim 46, wherein searching the database includes
2 searching with a subpattern of the first disulfide signature.

1 48. The computer program of claim 47, wherein the subpattern is generated by
2 calculating the disulfide signature that results when one or more disulfide bridges is removed
3 from the protein sequence corresponding to the first disulfide signature.

1 49. The computer program of claim 46, wherein at least one disulfide signature in
2 the database is associated with a sequence identifier.

1 50. The computer program of claim 46, wherein at least one disulfide signature in
2 the database is associated with a domain identifier.

1 51. The computer program of claim 46, further comprising clustering disulfide
2 signatures of the database.

1 52. The computer program of claim 51, wherein clustering includes grouping
2 disulfide signatures by number of disulfide bridges.

1 53. The computer program of claim 51, wherein clustering includes grouping
2 disulfide signatures by disulfide topology.

1 54. The computer program of claim 51, wherein clustering includes calculating a
2 measure of similarity between disulfide signatures and grouping based on the measure of
3 similarity.

1 55. A computer-readable data storage medium comprising a data storage material
2 encoded with a computer-readable database, the database comprising a plurality of disulfide
3 signatures, each disulfide signature of the database characteristic of a corresponding protein
4 sequence.

1 56. The data storage medium of claim 55, wherein each disulfide signature of the
2 database describes a disulfide topology of the corresponding protein sequence.

1 57. The data storage medium of claim 55, wherein each disulfide signature
2 includes the number of residues between a pair of cysteines joined by a disulfide bridge, and
3 the number of residues between the first cysteine of each disulfide bridge and the first
4 cysteine of the next disulfide bridge in the corresponding protein sequence.

1 58. The data storage medium of claim 57, wherein each disulfide signature
2 includes the number of residues between each pair of cysteines joined by a disulfide bridge,
3 and the number of residues between the first cysteine of each disulfide bridge and the first
4 cysteine of the next disulfide bridge in the corresponding protein sequence, for each disulfide
5 bridge in the corresponding protein sequence.

1 59. The data storage medium of claim 55, wherein at least one disulfide signature
2 in the database is associated with a sequence identifier.

1 60. The data storage medium of claim 55, wherein at least one disulfide signature
2 in the database is associated with a domain identifier.

1 61. The data storage medium of claim 55, wherein at least one disulfide signature
2 in the database is associated with a cluster identifier.

1 62. The data storage medium of claim 55, wherein the data storage material is
2 further encoded with a computer program comprising instructions for causing a computer
3 system to compare a first disulfide signature to a second disulfide signature, each disulfide
4 signature being characteristic of a corresponding protein sequence.

1 63. The data storage medium of claim 62, wherein comparing includes calculating
2 a measure of similarity between the first disulfide signature and the second disulfide
3 signature.

1 64. The data storage medium of claim 63, wherein comparing further includes
2 calculating a measure of statistical relevance for the measure of similarity between the first
3 disulfide signature and the second disulfide signature.

1 65. The data storage medium of claim 62, wherein comparing includes searching
2 the database.

1 66. The data storage medium of claim 65, searching the database includes
2 searching with a subpattern of the first disulfide signature.

1 67. The data storage medium of claim 66, wherein the subpattern is generated by
2 calculating the disulfide signature that results when one or more disulfide bridges is removed
3 from the protein sequence corresponding to the first disulfide signature.

1 68. A method of describing a protein sequence comprising generating a first
2 disulfide signature, the disulfide signature describing the cysteine spacing and disulfide
3 topology of first a protein sequence.

1 69. The method of claim 68, further comprising identifying a disulfide bridge by
2 protein sequence homology or protein structure homology.

1 70. The method of claim 68, further comprising generating a second disulfide
2 signature, the signature describing the cysteine spacing and disulfide topology of a second
3 protein sequence.

1 71. The method of claim 70, further comprising comparing the first disulfide
2 signature to a second disulfide signature.

1 72. The method of claim 71, wherein comparing includes calculating a measure of
2 similarity between the first disulfide signature and the second disulfide signature.

1 73. The method of claim 71, further comprising generating a database including
2 the first and second disulfide signatures.